Thinking Enough? Evaluating Advanced Large Language Models' Reasoning Algorithms in HEOR

MSR233

Tushar Srivastava, Shilpi Swami

ConnectHEOR Ltd., London, UK. Email: tushar.srivastava@connectheor.com

BACKGROUND

CONTEXT: Large Language Models (LLMs) are a type of foundation model, trained on massive datasets enabling them to recognize, summarize, and generate text, producing coherent and contextually relevant outputs. Health Economics and Outcomes Research (HEOR) researchers can use application programming interfaces to leverage public, standard-purpose LLMs like ChatGPT, Bard, Claude, or Cohere; medical LLMs like BioGPT-JSl or Hippocratic AI; or bespoke "closed" LLMs on proprietary document libraries to conduct systematic literature reviews, synthesize text, and generate first drafts and content variations for HEOR analysis.[1,2]

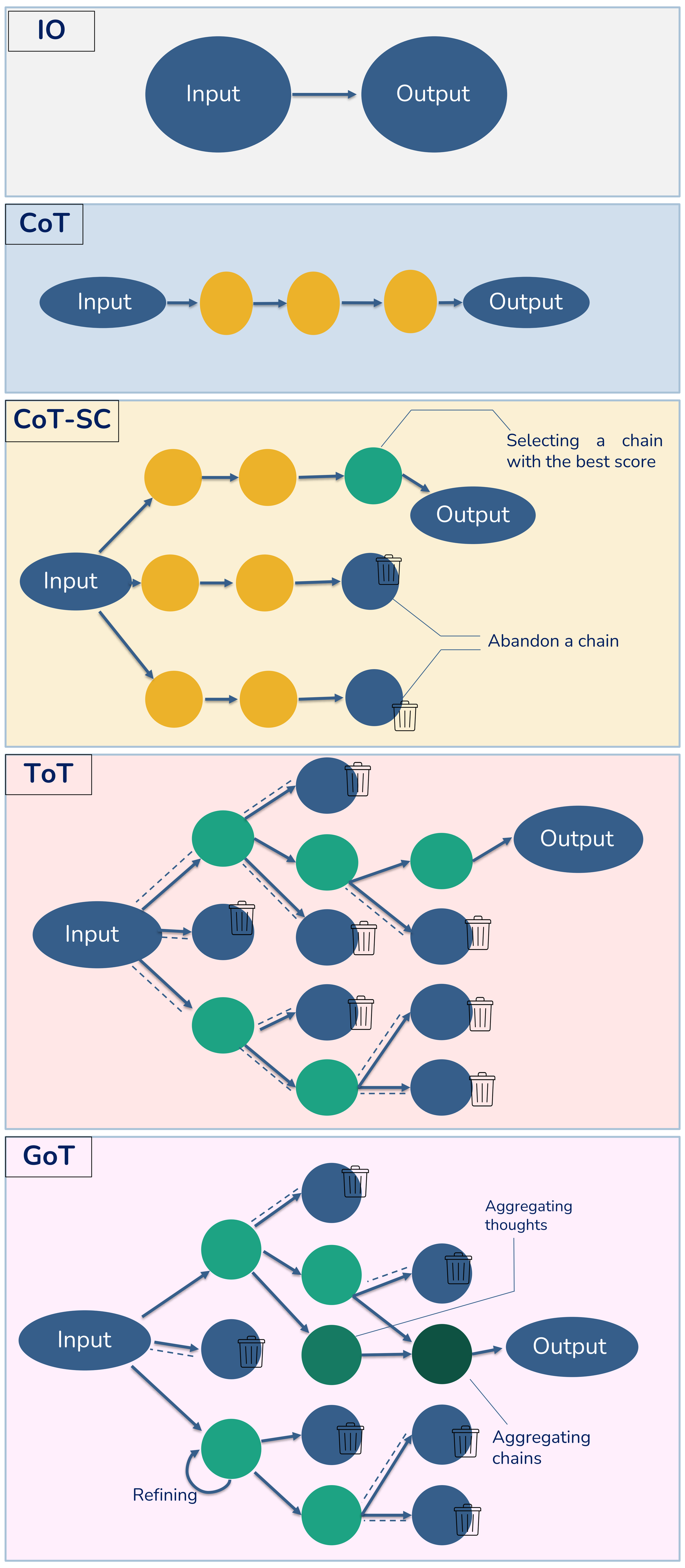
OBJECTIVES: To explore and evaluate various reasoning algorithms of LLM and their application in HEOR.

METHODS

- This study reviews and conducts a theoretical examination of existing advanced AI reasoning algorithms, assessing their potential for use in HEOR by enabling more complex, iterative, and nuanced decision-making processes. The evaluation focuses on how these algorithms can facilitate more dynamic and adaptive reasoning in HEOR.
- Reasoning algorithms in LLMs are designed to enable these models to perform complex reasoning tasks, similar to human thought processes. These algorithms allow LLMs to break down problems, analyze information, and generate coherent and logical responses.
- Prompt engineering modifies how this reasoning occurs, influencing the model to adopt certain reasoning strategies, work through problems iteratively, or apply specific constraints.
- Here are some key techniques: Basic Input Output (IO), Chain of Thoughts (CoT), Multiple Chains of Thoughts (CoT-SC), Tree of Thoughts (ToT), and Graph of Thoughts (GoT). [Figure 1]

INTEGRATION OF PROMPT ENGINEERING IN HEOR

Figure 1: Prompt engineering techniques



MECHANISM	USE CASE IN HEOR
• **Input Output Algorithm:** Simplest form of AI algorithms, directly generates output from input. • **Limitations:** Limited ability to incorporate broader contextual information, relies heavily on learned patterns, and lacks iterative improvement.	IO can support in general background queries on the disease, existing diagnostics and treatments, and scoping of HEOR studies.
• **Chain of Thoughts:** Includes a rationale for each example, breaking it down into manageable steps for clearer reasoning.[3] • **Limitations:** Linear reasoning, no scope for feedback looping or backtracking, and costly for large models.[3]	CoT is beneficial in HEOR when analyzing processes that unfold in a sequence, such as tracking the steps in a patient's journey through treatment/evaluating the stages of cost accumulation over time.
• **Chains of Thoughts-Self Consistency:** This approach asks the model same prompt multiple times, generating multiple responses. The final answer is determined by taking the majority results as a consensus. It improves the model's robustness to imperfect prompts, helps to gather rationales during reasoning tasks and maintain internal consistency.[4] • **Limitations:** Increases computational cost compared to standard CoT although a small number of paths can still yield significant performance improvements.[4]	CoT-SC allows to evaluate healthcare interventions by refining cost-effectiveness models through iterative reasoning, applying logic-based rules to assess cost-effectiveness, clinical efficacy, and safety. By generating multiple responses and selecting the consensus, it enhances the reliability of evaluations, especially when dealing with uncertain or complex data.
• **Tree of Thoughts:** It structures the AI's problem-solving process in a tree-like manner, branching out different pathways and scenarios. Enables an LLM to self-evaluate the progress through intermediate thoughts.[5] • **Limitations:** Can become computationally intensive with many branches. Bounded by a rigid tree structure. Each query to online LLM APIs such as GPT-4 not only incurs a monetary expense but also contributes to latency.[5]	ToT enables to map out and explore diverse clinical pathways a patient might experience. In a clinical setting, this allows for the consideration of different patient responses, treatment side effects, and progression scenarios. By visualizing these branching pathways, clinicians can better anticipate potential outcomes based on individual patient characteristics, supporting personalized treatment plans.
• **Graph of Thoughts:** It has the ability to model the information generated by an LLM as an arbitrary graph, where units of information ("LLM thoughts") are vertices, and edges correspond to dependencies between these vertices. It enables AI to consider multiple pathways and relationships simultaneously.[6] • **Limitations**: This approach might incur higher computational costs compared to linear reasoning algorithms. Time intensive.[6]	GoT allows to builds models for comorbid conditions, capturing interactions between treatments and outcomes in a knowledge graph. Each node (vertex) in the graph could represent a specific health state, treatment option, or outcome, while the edges represent transitions between these states based on different interventions, could be useful in synthesizing research evidence, especially in terms of visualizing how different factors (clinical, economic, quality of life) interact or influence each other.

CONCLUSION

Advanced AI reasoning algorithms, beyond the basic IO approach, hold significant potential to offer deeper analysis and more refined reasoning outcomes in HEOR. The shift from traditional linear reasoning to more sophisticated, network-based models allows for the exploration of complex decision avenues, greatly enhancing analytical capabilities and accuracy.

While LLMs hold significant promise in advancing HEOR practices, it is crucial to continuously assess their reasoning capabilities and adapt them to meet the evolving needs of healthcare research. As these models continue to develop, their potential to revolutionize HEOR remains substantial, provided that their implementation is done with caution and a strong focus on practical, real-world applicability and ethical usage.

FUTURE SCOPE

- ✓ **Enhanced causal reasoning:** Future advancements could improve the ability of LLMs to understand and predict the causal effects of healthcare interventions, leading to more accurate and reliable outcomes.
- ✓ **Integration with real-world data:** Combining these reasoning models with real-world data sources can enhance the accuracy and applicability of HEOR analyses.
- ✓ **Personalized medicine:** These models can support the development of personalized treatment plans by considering a wide range of variables and potential outcomes, leading to more tailored and effective healthcare interventions.

References.

1. Fleurence RL, et al. Generative AI in Health Economics and Outcomes Research: A Taxonomy of Key Definitions and Emerging Applications - an ISPOR Working Group Report, 2024.
2. Liu P, Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, 2021
3. Wei J, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, 2023
4. Wang X ,Self-consistency improves chain of thought reasoning in language models, 2023.
5. Yao S, Tree of Thoughts: Deliberate Problem Solving with Large Language Models, 2023
6. Besta M, Graph of Thoughts: Solving Elaborate Problems with Large Language Models.

Acknowledgments

The authors acknowledge Shashwat Gaur and Hanan Irfan (ConnectHEOR, Delhi, India) for developing and designing the content of the poster. ***Manuscript in writing.***